ROBUST DESIGN USING

SEQUENTIAL COMPUTER EXPERIMENTS

A Thesis

by

ABHISHEK GUPTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2004

Major Subject: Industrial Engineering

ROBUST DESIGN USING

SEQUENTIAL COMPUTER EXPERIMENTS

A Thesis

by

ABHISHEK GUPTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Yu Ding (Chair of Committee)	Alberto Garcia-Diaz (Member)
Larry J. Ringer (Member)	Mark L. Spearman (Head of Department)

May 2004

Major Subject: Industrial Engineering

ABSTRACT

Robust Design Using Sequential Computer Experiments. (May 2004)
Abhishek Gupta, B.Tech. , Indian Institute of Technology, Delhi, India
Chair of Advisory Committee: Dr. Yu Ding

Modern engineering design tends to use computer simulations such as Finite Element Analysis (FEA) to replace physical experiments when evaluating a quality response, e.g., the stress level in a phone packaging process. The use of computer models has certain advantages over running physical experiments, such as being cost effective, easy to try out different design alternatives, and having greater impact on product design. However, due to the complexity of FEA codes, it could be computationally expensive to calculate the quality response function over a large number of combinations of design and environmental factors. Traditional experimental design and response surface methodology, which were developed for physical experiments with the presence of random errors, are not very effective in dealing with deterministic FEA simulation outputs. In this thesis, we will utilize a spatial statistical method (i.e., Kriging model) for analyzing deterministic computer simulation-based experiments. Subsequently, we will devise a sequential strategy, which allows us to explore the whole response surface in an efficient way. The overall number of computer experiments will be remarkably reduced compared with the traditional response surface methodology. The proposed methodology is illustrated using an electronic packaging example.

To my parents

ACKNOWLEDGEMENTS

First and foremost I thank my thesis committee - Dr. Yu Ding, Dr. Alberto-Garcia Diaz and Dr. Larry J. Ringer - for their time and thoughtfulness. My advisor, Dr. Yu Ding sets an example of an engaged teacher and avid scholar who serves as a model for his students. From the beginning, he has been a source of constant encouragement and inspiration. Through his mentorship, I feel I am a better researcher.

My sincere thanks go to Dr. Leon Xu and Mr. Tommi Reinikainen of Nokia Inc., for their valuable suggestions during the project. I hope my work is of use to them.

I probably would not be writing this if it were not for the constant support from my family. Thank you for your endless love and patience throughout my career.

I am also thankful to all of my friends for their support and making life at Texas A&M very enjoyable.

"*The learning and knowledge that we have, is, at the most, but little compared with that of which we are ignorant.*" - Plato

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

With the increasingly competitive environment, there is a need to design and produce more technically complex products, with assured product reliability in shorter time than the current practice. In the mobile phones market, this translates into better versatility, portability and visual and ergonomic appeal amongst other features. This in turn calls for continuous improvement in electronic packaging.

A potential problem with the electronic packaging for mobile phones could be caused by the failure of the solder joints under thermal and mechanical loading. Figure 1 shows the bending process map, where the Von Mises stress generated in the joints is used as a key parameter to assess solder reliability under bending. Hence, the objective is to minimize the maximum stress in solder joints with respect to various uncontrollable environmental noises or disturbances.

Robust optimal design is to design a system (a product or a process) insensitive to noise/disturbance existing under normal production and usage conditions. We analyze the functional response surface so that the settings of controllable design variables are optimized to minimize the maximum stress and reduce the sensitivity of response to uncontrollable environmental noises.

The first step for robust design is to classify the input factors as controllable factors or design variables, $\mathbf{x}_c$, and noise factors or environmental variables, $\mathbf{x}_e$. We suppose the domain of y is χ which is a fixed subset of $\Re^d$ and $\mathbf{x}_c, \mathbf{x}_e \in \chi$. The process response y can be generally expressed as $y = f(\mathbf{x}_c, \mathbf{x}_e)$. A robust design can be achieved by minimizing the expectation of y, $L(\mathbf{x}_e)$, with respect to the distribution of noise factor, $g(\mathbf{x}_e)$, i.e.,

---

This thesis follows the style and format of ASME Journal of Manufacturing Science and Engineering.

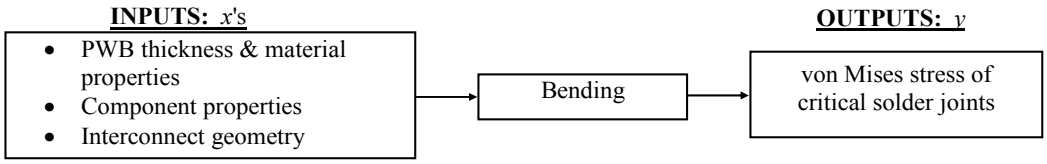


Fig. 1 Bending process map

$$\text{Minimize}_{\{\mathbf{x}_c\}} \quad L(\mathbf{x}_c) = \int_{\mathbf{x}_e \in \chi} f(\mathbf{x}_c, \mathbf{x}_e) g(\mathbf{x}_e) d\mathbf{x}_e \tag{1}$$

In this thesis, we discuss a robust design methodology applicable to electronic packaging. Our approach involves knowledge of $f(\mathbf{x}_c,\mathbf{x}_e)$. Given the complexity of the packaging process, it is impossible in most practical cases to know $f(\mathbf{x}_c,\mathbf{x}_e)$ analytically. Thus, the above integral has to be evaluated numerically. In fact, Finite Element Analysis (FEA) models are widely used in evaluating electronic packaging mechanics with a lot of success [1-3]. When using a FEA model, equation (1) could be numerically computed as

$$\text{Minimize}_{\{\mathbf{x}_c\}} \quad L(\mathbf{x}_c) = \int_{\mathbf{x}_e \in \chi} f_{\text{FEA}}(\mathbf{x}_c, \mathbf{x}_e) g(\mathbf{x}_e) d\mathbf{x}_e \tag{2}$$

where $f_{\text{FEA}}(\mathbf{x}_c, \mathbf{x}_e)$ is the FEA model of the true process $f(\mathbf{x}_c, \mathbf{x}_e)$.

To illustrate, consider a generic model of a 2nd level packaging (electronic package to board) design under bending loading. The objective is to assure electronic packaging robustness and bending reliability of solder joints for minimum von Mises stresses by optimizing the design variables such as geometric dimensions and material properties of the model. Figure 2 shows the Chip Scale Package (CSP)-Printed Wiring Board (PWB) model used in this study. The stress conditions under bending load in the PWB is modeled and calculated using a commercial FEA software, *ANSYS* [4] .

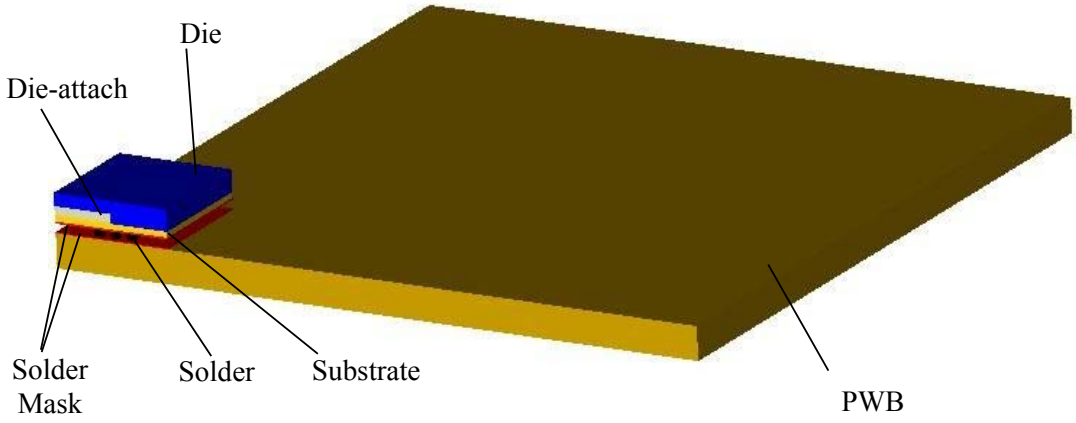


Fig. 2 The CSP-PWB model

In this study, engineers from our collaborating company have already identified four important design parameters related to the material properties of the PWB. They are: PWB in-plane Young's modulus (x_1), component substrate in-plane Young's modulus (x_2), die attach Young's modulus (x_3), and molding compound Young's modulus (x_4). Table 1 provides the allowable parameter ranges defining the design region. Obviously, for this particular example, a robust design is to find the robust parameter setting of material properties that minimize the maximum Von Mises stress (y) in the solder joints.

The use of computer models has certain advantages over running physical experiments such as being cost effective, easy to try out design alternatives, and thus having greater impact on product design. However, due to the complexity of the FEA codes, it could be computationally expensive to calculate $f_{\text{FEA}}(\mathbf{x}_c,\mathbf{x}_e)$ over a large number of combinations of $\mathbf{x}_c$ and $\mathbf{x}_e$. Therefore, the key issue for realizing a robust design is finding a simple and computationally economical model to approximate $f_{\text{FEA}}(\mathbf{x}_c,\mathbf{x}_e)$.

The model fitting methods will be very different for deterministic computer simulation outputs and physical experimental data. The difference in data analysis strategies for the physical and computer experiments is shown in Fig. 3. In physical

Table 1 Design space for the material property parameters (Unit: GPa)

	PWB in-plane Young's modulus (x_1)	Substrate in-plane Young's modulus (x_2)	Die attach Young's modulus (x_3)	Molding compound Young's modulus (x_4)
High	35	36	5	30
Low	20	12	1	15

experiments, random errors are always present and hence the experimenters are encouraged to take replicates. The prediction from such replicated data is done through the fitting of trend lines, e.g. from linear regression. In computer experiments, linear regression can still be used to get a trend line. However, it does not hold a clear meaning in the absence of random errors. Recently spatial statistical methods like the Kriging model [5] have been used to analyze the deterministic computer data. Kriging model is an interpolative model, which will fit a predictor to pass through all the observed points because there is no uncertainty involved in the observed value at a particular parametric level (refer to Fig. 3 (b)) . On the other hand, we usually do not employ such an interpolative approach in the case of a physical experiment because each observation is associated with uncertainty and the interpolative approach will cause over-fitting and thus poor prediction. The Kriging predictor in computer experiment, as it passes through all the observed points, is more complicated than a first or second order polynomial. For other combinations of parameter levels where an observation (i.e., a computer output) is not available, the predicted value will be obtained from interpolation using the Kriging model developed from the observed points.

In this thesis, our focus is to obtain the functional response function for a PWB product based on FEA simulations and a Kriging model. Once a computationally economical surrogate of the response function $f_{\mathrm{FEA}}(\mathbf{x}_c,\mathbf{x}_e)$ is known, it would be straightforward to numerically optimize equation (2) for a robust design solution, based on identification of noise factors and assumptions on their distributions.

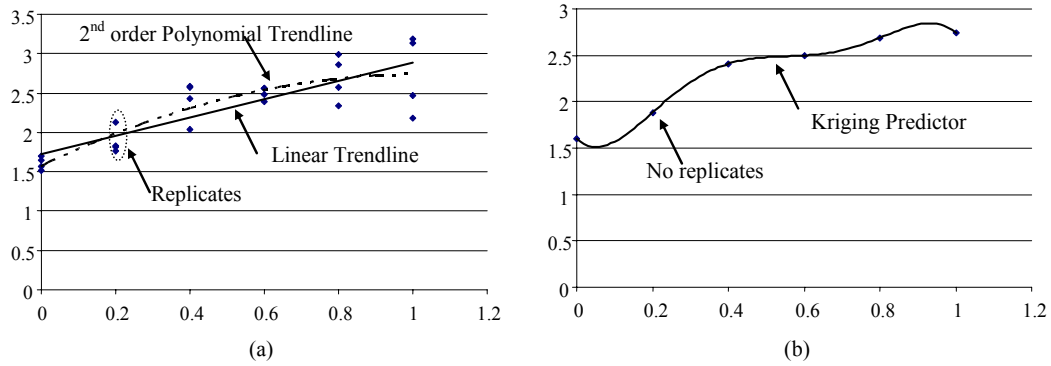


Fig. 3 Difference between (a) physical and (b) computer experiments

As is the case with most modeling approaches, an accurate Kriging model over the entire design space would require a large number of computer FEA experiments. We will devise a sequential strategy to address the issue of how to reduce the number of computer experiments. The basic idea is as follows. We are oftentimes more concerned with a sub-region or sub-regions of the design space which contain the optimal parameter combinations, which suggests that collecting data with the same density throughout the design space may be unnecessary. In the sequential strategy, we choose in the initial step to use a small number of computer experiments, covering the whole design space. Subsequently we keep zooming into a smaller sub-region of interest based on the predicted values and prediction uncertainty of the Kriging predictor calculated from the previously steps. More design points will be only added to the small sub-region so that we can have additional information to fit a more accurate model until a reasonably accurate optimum is found. This sequential design process will be applied to the aforementioned electronic packaging design and its benefit will be demonstrated by a comparison with a traditional RSM design.

This thesis includes five chapters. Following the introduction chapter, Chapter II presents an overview of Kriging model. Chapter III presents the details of the sequential

strategy and applies the sequential methodology to the electronic packaging design. Chapter IV provides a comparison of the sequential Kriging strategy with the traditional RSM methodology. Finally, Chapter V concludes the thesis.

CHAPTER II

REVIEW OF KRIGING METHOD FOR COMPUTER EXPERIMENTS

Sacks et al. [5,6] and Welch et al. [7] first proposed the use of Kriging models for analyzing computer experiments. The idea of Kriging for computer experiments was extended from geo-statistics [5,8]. Since we are interested in obtaining the response surface of the function, our case is very similar to the problem of establishing correlations among spatially distributed locations. Spatial statistics models such kind of spatial correlation using a stochastic modeling approach that views the spatial response function as a realization of a random field (also sometimes called random function or stochastic process). Assume that $y(\mathbf{x})$ is a response function of a d-dimensional vector of inputs $\mathbf{x}$ over domain χ. The notation $Y(\cdot)$ is used to distinguish the random function from its realizations $y(\cdot)$. $Y(\mathbf{x})$ is most commonly given by

$$Y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) \tag{3}$$

where $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \dots \quad \beta_k]^T$ is the vector of unknown regression coefficients, $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \quad f_2(\mathbf{x}) \quad \cdots \quad f_k(\mathbf{x})]^T$ is a vector of known regression functions, and $Z(\mathbf{x})$ is a zero mean stationary Gaussian random field over χ. The intuition behind the model in equation (3) is that while the regression portion of the model approximates the response surface "globally", the local deviations are captured by the $Z(\mathbf{x})$ component. In other words, $Z(\mathbf{x})$ tries to capture the systematic departure from the global regression part. For this goal, $Z(\mathbf{x})$ is assumed to have covariance

$$\sigma^2 R(\mathbf{w}, \mathbf{x}) \tag{4}$$

between $Z(\mathbf{w})$ and $Z(\mathbf{x})$ at two vector-valued inputs $\mathbf{w}$ and $\mathbf{x}$ in χ, where σ^2 is the process variance and $R(\mathbf{w}, \mathbf{x})$ is the correlation function. In order to make valid statistical inference using a single realization $y(\mathbf{x})$ of the random process $Y(\cdot)$, we need

the random process to be ergodic and stationary [9]. For this reason, Kriging models as used in this thesis assume that $Z(\mathbf{x})$ is second-order covariance stationary, i.e., the mean is constant and $\mathrm{Cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_1 + \mathbf{h})\}$ is independent of $\mathbf{x}$ and depends only on the distance $\mathbf{h}$. This is in contrast to the covariance structure of $Z(\mathbf{x})$ in the traditional RSM for physical experiments, where the covariance matrix is assumed to be $\sigma^2\mathbf{I}$. The $\sigma^2\mathbf{I}$ covariance is introduced to represent the randomness associated with the replication at given design inputs. It, however, fails to capture the spatial correlation between two design inputs. Therefore, the new covariance structure of $Z(\mathbf{x})$ provides the Kriging predictor interpolative capability, as desired for analyzing deterministic computer experiments.

In order to use the Kriging predictor in a practical setting, researchers parameterize the correlation matrix. The most popular family of correlation models in the computer experiments literature is the power exponential correlation family [10]. Product of stationary one-dimensional correlations gives us,

$$R(\mathbf{w}, \mathbf{x}) = \prod_{j=1}^{d} \exp(-\theta_j \left| w_j - x_j \right|^{p_j}) \tag{5}$$

where $\theta_j \geq 0$, $0 \leq p_j \leq 2$ and $\mathbf{w}, \mathbf{x} \in \chi$ with w_j being the j^{th} component of $\mathbf{w}$ and x_j being the j^{th} component of $\mathbf{x}$. In particular we will use $p_j = 2$, because it is infinitely differentiable at zero. To realize the Kriging model, we need to estimate unknown parameters σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta} = (\theta_1, \ldots \theta_d)$ from the computer simulation data (in our case we take $p_j = 2$ for all j's).

We first introduce the following notations. The n design points are denoted by $\mathbf{D} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and the observed response vector corresponding to the design points is $\mathbf{y}_D = \left[y(\mathbf{x}_1) \quad \cdots \quad y(\mathbf{x}_n) \right]^T$. Define $\mathbf{R}_D$ as the $n \times n$ matrix of correlations between Z's at the design points, where $\mathbf{R}_{D(i,j)} = R(\mathbf{x}_i, \mathbf{x}_j)$ for $1 \leq i, j, \leq n$. Denote by $\mathbf{F} = \left[\mathbf{f}^T(\mathbf{x}_1) \quad \cdots \quad \mathbf{f}^T(\mathbf{x}_n) \right]^T$ the $n \times k$ design matrix.

We will employ the maximum likelihood estimation to estimate σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. Under

normality assumption, the log-likelihood function of σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta} \mid \mathbf{y}_D) = -\frac{1}{2}\left[n\log\sigma^2 + \log(\det(\mathbf{R}_D)) + (\mathbf{y}_D - \mathbf{F}\boldsymbol{\beta})^T \mathbf{R}_D^{-1}(\mathbf{y}_D - \mathbf{F}\boldsymbol{\beta})/\sigma^2\right] \quad (6)$$

The above equation is seldom directly used for obtaining the maximum likelihood estimates (MLE) of σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. It is usually simplified as follows. Assuming that $\boldsymbol{\theta}$ is known, we can get the MLE of $\boldsymbol{\beta}$ and σ^2 as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{F}^T\mathbf{R}_D^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{R}_D^{-1}\mathbf{Y}_D \text{ and } \hat{\sigma}^2 = \frac{1}{n}(\mathbf{y}_D - \mathbf{F}\hat{\boldsymbol{\beta}})^T\mathbf{R}_D^{-1}(\mathbf{y}_D - \mathbf{F}\hat{\boldsymbol{\beta}}), \quad (7)$$

respectively. Substituting them in equation (6), we can get a simplified log-likelihood function with $\boldsymbol{\theta}$ as the only unknown parameter vector,

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}_D) = -\frac{1}{2}\left[n\log\hat{\sigma}^2 + \log(\det(\mathbf{R}_D)) + n\right]. \quad (8)$$

This equation is usually solved by numerical algorithms such as the Newton-Raphson [11], Nelder-Mead's simplex search [12] or the E-M method [11]. In our implementation, we used the simplex search, which is readily available in MATLAB, to solve equation (8) for $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$. Then, substitute $\hat{\boldsymbol{\theta}}$ back to equation (7) for $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$. Given $\hat{\sigma}^2$ and $\hat{\boldsymbol{\theta}}$, the covariance matrix for $\{Z(x_i)\}_{i=1}^n$ is $\hat{\sigma}^2\hat{\mathbf{R}}_D$, where $\hat{\mathbf{R}}_D$ is the correlation matrix with $\hat{\theta}_j$'s in the place of θ_j's.

In computer experiments, we want to predict y_0 at an untried input $\mathbf{x}_0$, based on the observed data $\mathbf{y}_D$. Because we are using Gaussian random fields, both observed and unobserved values can be pooled together in a multivariate Gaussian vector. Thus the joint distribution of Y_0 and $\mathbf{Y}_D$ is

$$\begin{pmatrix} Y_0 \\ \mathbf{Y}_D \end{pmatrix} \sim N_{1+n}\left[\begin{pmatrix} \mathbf{f}_0^T \\ \mathbf{F} \end{pmatrix}\boldsymbol{\beta}, \sigma^2\begin{pmatrix} 1 & \mathbf{r}_0^T \\ \mathbf{r}_0 & \mathbf{R}_D \end{pmatrix}\right] \quad (9)$$

where $\mathbf{f}_0 = \mathbf{f}(\mathbf{x}_0)$ and $\mathbf{r}_0^T = \left(R(\mathbf{x}_0 - \mathbf{x}_1),\ldots,R(\mathbf{x}_0 - \mathbf{x}_n)\right)$ is the $n\times 1$ vector of correlations of Z's at the design points and the untried input $\mathbf{x}_0$. Substituting the σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ with their MLEs and assuming $\hat{\mathbf{R}}_D$ as positive definite, Santner et al. [10] derives the best

empirical mean square error (EMSE) predictor of y_0 as

$$\hat{Y}_0 = \mathbf{f}_0^T \hat{\boldsymbol{\beta}} + \hat{\mathbf{r}}_0^T \hat{\mathbf{R}}_D^{-1} (\mathbf{Y}_D - \mathbf{F}\hat{\boldsymbol{\beta}}) \tag{10}$$

The model can be thought to be a sum of the generalized least squares predictor $\mathbf{f}_0^T \hat{\boldsymbol{\beta}}$ and the second term $\hat{\mathbf{r}}_0^T \hat{\mathbf{R}}_D^{-1} (\mathbf{y}_D - \mathbf{F}\hat{\boldsymbol{\beta}})$ which is the smooth of the residuals $(\mathbf{y}_D - \mathbf{F}\hat{\boldsymbol{\beta}})$. The mean square error (MSE) for the predictor can be calculated as

$$\mathrm{MSE}[\hat{Y}_0] = \sigma^2 \left[1 - (\mathbf{f}^T(\mathbf{x}_0) \quad \hat{\mathbf{r}}_0^T(\mathbf{x}_0)) \begin{pmatrix} \mathbf{0} & \mathbf{F}^T \\ \mathbf{F} & \hat{\mathbf{R}}_D \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{f}(\mathbf{x}_0) \\ \hat{\mathbf{r}}_0(\mathbf{x}_0) \end{pmatrix} \right] \tag{11}$$

For details of the above derivations, please refer to [5] or [10]. The subsequent chapter will develop a sequential strategy, using a Kriging predictor and evaluate its uncertainty recursively.

CHAPTER III

THE SEQUENTIAL STRATEGY

3.1 Overview of the Sequential Strategy

Suppose we are searching for the maximal value over a response surface as shown in Fig. 4, where multiple peaks (or valleys for a minimization problem) are present. If we employ a traditional RSM, its success in finding the global maximum depends on the starting point. That is to say, the traditional RSM will be easily entrapped in a local maximum (or minimum) by ascending (or descending) the wrong hill (or valley). This brings us to the important issue of the lack of prior knowledge of the response surface. Before the first experiment, we usually have no information about the response surface and the current operating condition is often chosen as the starting point. Such a starting point has no guarantee to take us to the optimal setting.

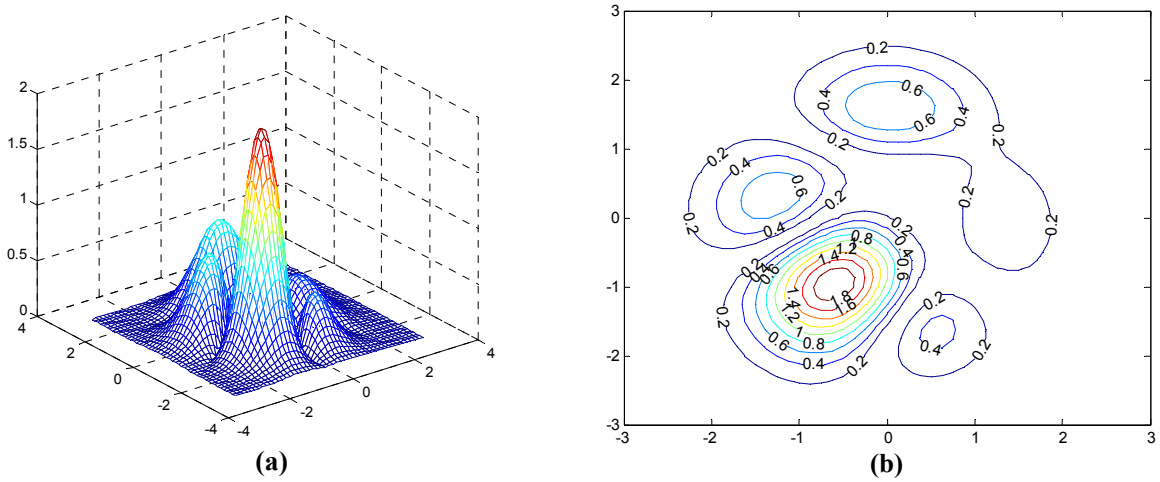


Fig. 4 Illustration of a complicated response surface

The Kriging model does not depend on any starting point. It tries to have an approximating envelope passing through a sample of design points. Obviously in order not to miss the optimal areas, we should spread design points as evenly as possible, i.e., we need to employ a space filling design. Suppose to model the above surface we choose a 16-point Latin hypercube sampling design (described in detail later), then using a Kriging model we can fit a response surface. The contour plot of the response surface generated by the Kriging model is shown in Fig. 5 (a). This fitted response offers a reasonably good representation of the overall surface and captures the peak area.

Intuitively it appears that if we add more design points to a sub-region and then revise our Kriging model based on the modified design, it should be able to represent the true surface better than the previous one. This simple idea motivates our sequential experimental strategy. The major decision is to select an appropriate zooming-in sub-region. There are two primary considerations. One is to zoom in to the areas of interest, i.e., a sub-region where we find large values of objective function (in a maximization problem). Another is to consider the uncertainty associated with the Kriging prediction and add points to the sub-regions with high uncertainty values. The second consideration is to reduce the likelihood that an initial rough estimation may miss potential optimum areas – for example, other peaks in Fig. 5 (a) are not very clearly identified. One can certainly choose to add points at sub-regions to satisfy both considerations.

For the example in Fig. 4, we chose to use the first criterion. Figure 5 (a) shows the sub-region selected and Fig. 5 (b) shows the contour plot after we have added 9 points in the zoom-in sub-region. Together with the initial 16 points, the 25-point design makes the prediction in that sub-region more accurate. Actually, the region outside the sub-region is also benefited, as can be seen by the clearer boundaries that have started to form for the other peak area. The procedure can be iteratively carried out until we see no significant improvement.

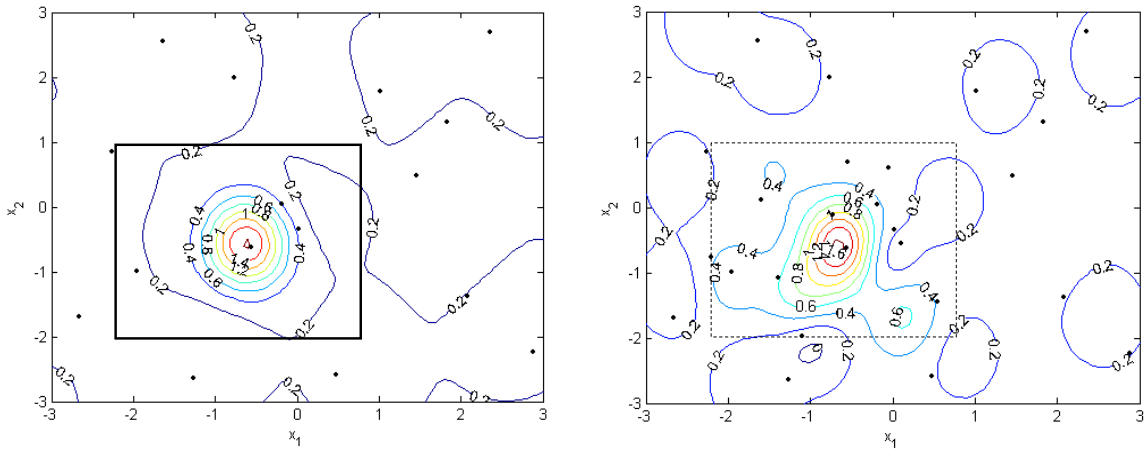


Fig. 5 Illustration of the sequential strategy

The general idea of sequential experiments has also been applied to other applications, such as in integrated circuit design by Bernardo et al. [13] and artificial joint design by Chang et al. [14]. Although the general idea bears certain similarity, there are many subtle decisions to make during this process because of different applications under investigation. The details of our sequential strategy will be illustrated using the electronic packaging example.

3.2 Initial Step

The first step in the sequential strategy is very important as all our subsequent decisions are based on the model we get after this step. Thus, we want to ensure that this model is as good a representation of the response surface as possible. In this step we need to

1. Choose a Kriging model, i.e., the regression polynomial and the covariance structure, which will model the functional response.
2. Choose an experimental design to obtain a sample of data points from the design space.

3. Compute the estimators for the Kriging model and develop the predictor.

The regression polynomial and the covariance structure for the Kriging model are often chosen based on the experimenter's experience or prior knowledge about the process. In our electronic packaging design problem, we choose a constant for the global regression part in equation (3). The constant is selected in lieu of any polynomial equation just as a starting expression. Our aim in this procedure is to keep the number of unknowns as low as possible. Based on the first step analysis, we will have better sense if we should add polynomial terms (the first order or the second order). If we find that satisfactory results are not obtained from just a constant regression component then we will have to add the appropriate polynomial terms. In fact, we will observe from the latter analysis, the Kriging model with a constant polynomial term is rather capable of capturing a complicated surface topology with deterministic trends. Thus for this electronic packaging problem, we have $f_1(\mathbf{x}) = 1$ and β_1.

As for the covariance function we choose the one given in equation (5) and for reasons discussed in Chapter II we assume $p_j = 2$. Thus for the electronic packaging model we have 6 unknown parameters, namely σ^2, $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3, \theta_4)$, and β_1.

Once the approximating model is chosen, we need an experimental design so that the computer experiments are run at these points and data are generated. In comparison to the design of physical experiments, where the data points are discrete factorial combinations and replicates are used to account for the random variation, the design for deterministic computer experiments should be spread throughout the design space without any replications. Such designs are known as the space-filling design [10]. Introduced by McKay et al. [15], Latin Hypercube Sampling (LHS) design is one of the most commonly used space-filling designs. An LHS design yielding n design points involves stratifying the design space into n equal probability intervals for each dimension, randomly selecting a point in each stratum and combining them to get a design point.

One critical question for time-expensive computer experiments is how to select the number of design points. We would always like to take lesser number of design points

but keeping in mind not to affect the fidelity of the model appreciably. The least number of design points we need to choose is the number of unknown parameters; we will otherwise have a singular problem. Bernardo et al. [13] suggested as a rule of the thumb choosing number of design points to be 3 times the number of unknown parameters for superior prediction performance. This means for a case like ours, where we have 6 unknown parameters we need at least 6 but could have as many as 18 points, depending on the prediction requirement. Bearing in the mind that we will have subsequent experiments to revise the first step model yet the first step should provide a good global view, we recommend choosing the number of design points in the initial step to be 1.5 to 2 times the number of unknown parameters in a general sequential experiment.

In this particular design, we choose 1.5 times the number of unknown parameters. This implies that we have 9 design points in the initial step and we choose a 9×4 (d=4) LHS design, where the design matrix is shown in the Appendix. Simulations corresponding to this design were carried out in *ANSYS* and the corresponding simulated Von Moses values (y) are included in the Appendix as well. A Kriging model was

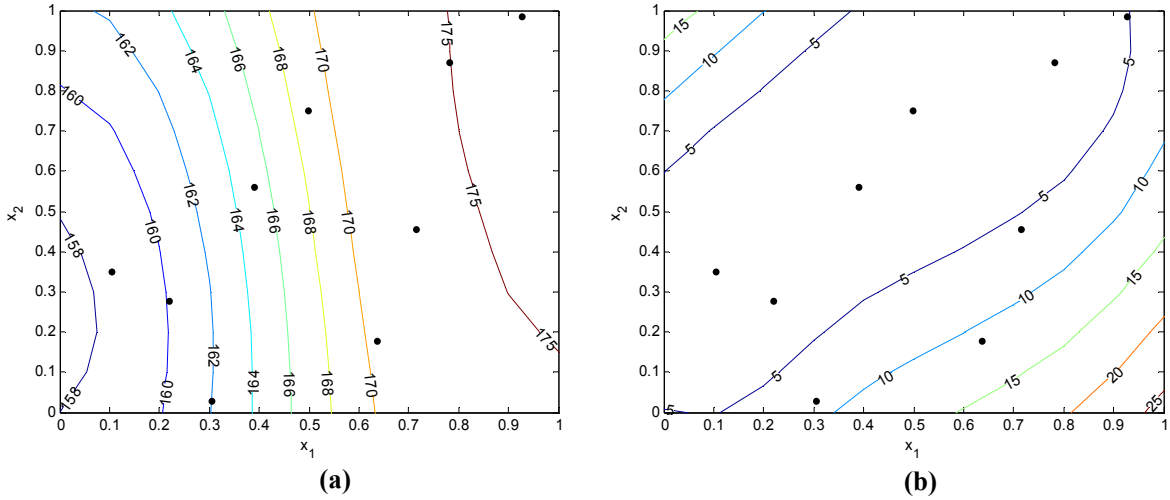


Fig. 6 Contour plots: (a) response surface and (b) MSE value for the 9-point model after the initial stage

Table 2 Results after the initial step

x_1	x_2	x_3	x_4	y	$\hat{y}$
20.0	16.8	3.0	16.5	154.44	157.53

constructed after calculating $\hat{\sigma}^2$, $\hat{\boldsymbol{\theta}}$, and $\hat{\beta}$ as $\hat{\sigma}^2 = 37.81$, $\hat{\boldsymbol{\theta}} =$ (2.5341, 0.6286, 0.4502, 1.0967), and $\hat{\beta} = 170.47$.

Finally, the Kriging model is used to generate predicted values over the design space. We can also use equation (11) to calculate the uncertainty associated with the model. In order to have a rough idea of what the response surface looks like, we display the contour plot of the response and its MSE value in Fig. 6 using two input variables x_1 and x_2 (other variable pairs can be used in the same spirit), where the black dots indicate the locations of nine design points.

We use this Kriging predictive model to explore the 4-dimensional design surface. Optimization routines such as simplex search are often employed to find the optimal region. In order to avoid to be entrapped into local optima, random search based algorithms (GA or SA) are also used. In our case, since we have a relatively low dimension, we simply choose to generate predicted values over a lattice of the design space for obtaining the minimum. The predicted and actual minima together with the corresponding design parameter values are given in Table 2.

3.3 Subsequent Steps

Naturally, we will add more design points in the following steps to improve our model prediction and find the minimal response value. The key question is where we should add additional design points. In this example, we will use both zoom-in criteria: (i) sampling more points over a region of interest; (ii) add extra points in a region of high uncertainty. Please note that selection of zoom-in region is subjective in nature, depending on design objectives and designers' understanding of underlying physics. We recommend using graphic tools such as main effect plots or response contour plots to

facilitate designers to select a region of interest.

3.3.1. Reduce the overall uncertainty level -- the 2nd stage

The next step is to reduce the prediction uncertainty because existence of a high uncertainty will mislead us to a region where the true response value is not small at all and thus taint our efforts in finding the minimum region. MSE contour plots for pair-wise design variables are shown in Fig. 7 (a)-(f). Given four design variables, we have six such plots. We can certainly observe some regions with high uncertainty. For instance, given the true response around 170, an MSE of 17 accounts roughly for 10% of the response value. We thus identify zoom-in sub-regions with a 10% error, which translates to a boundary with an MSE value between 15 and 20. Following this thought and also utilizing our graphic plot, we marked (with a thicker dash line) on each contour plot our zoom-in sub-region, e.g., it is $0.8 \le x_1 \le 1.0$ and $0 \le x_2 \le 0.3$ on Fig. 7 (a). Similar procedure applies to other plots in Fig. 7 and we eventually take the union of all the sub-regions as the region to be zoomed in.

$$0.8 \le x_1 \le 1.0; \quad 0 \le x_2 \le 1.0; \quad 0 \le x_3 \le 1.0; \quad 0 \le x_4 \le 1.0 \tag{12}$$

We then choose a 6-point (which is the number of unknown parameters) LHS design over the zoom-in region and concatenate the data points to our previous design to get a modified 15-point design. Please refer to the Appendix for the design matrix.

When adding extra design points over the zoom-in regions, we should also make sure the design points spread out. Therefore, a LHS design procedure will be followed with a small modification. Suppose that we need add k points in a region that already has p points, we will divide this region into $k+p$ bins and select points from the bins that are empty. This procedure ensures a more uniform space filling in the sub-region as compared to randomly take points from another LHS and just adding it to the original design.

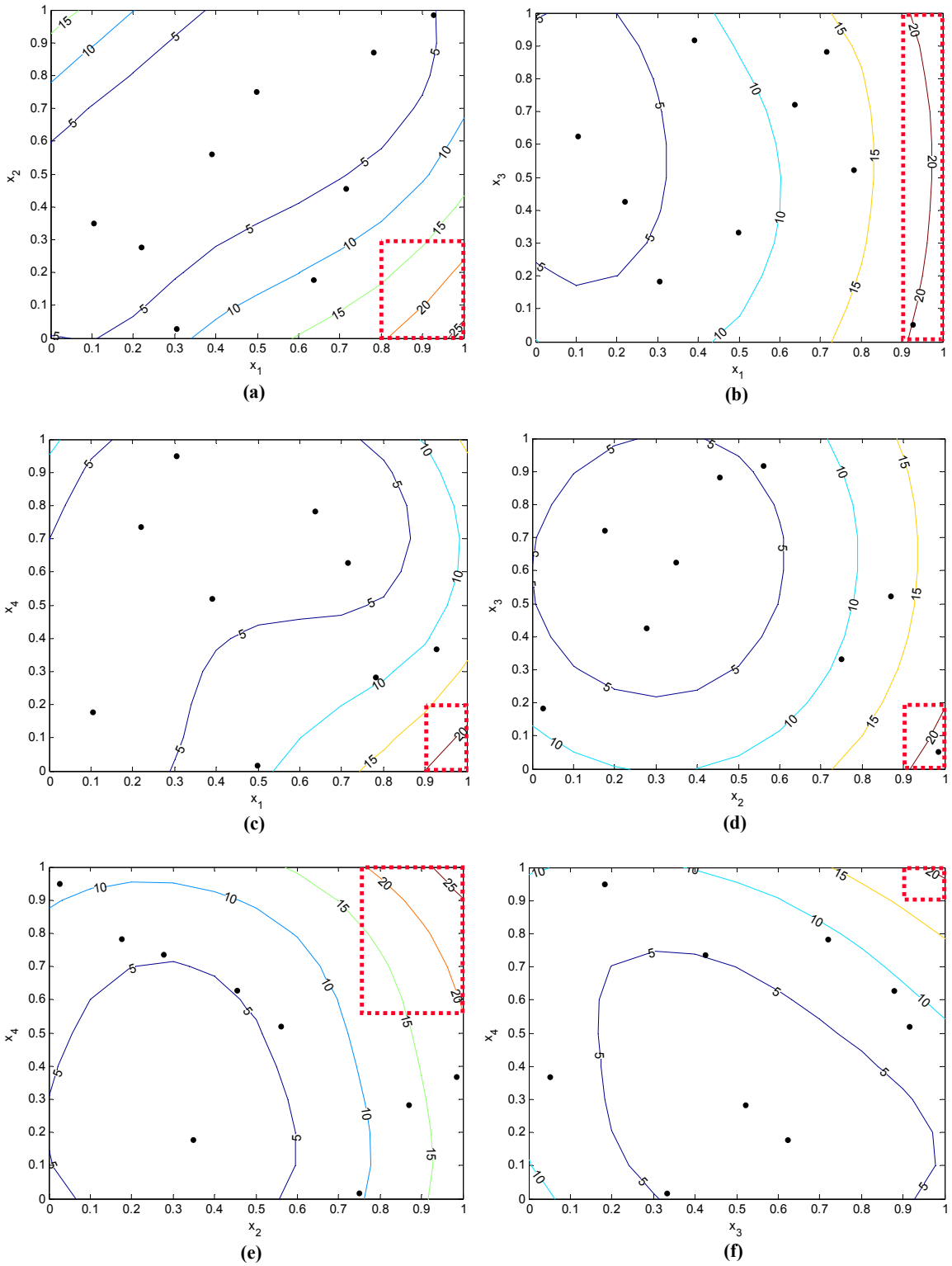


Fig. 7 MSE plots for pair-wise design variables after the initial stage

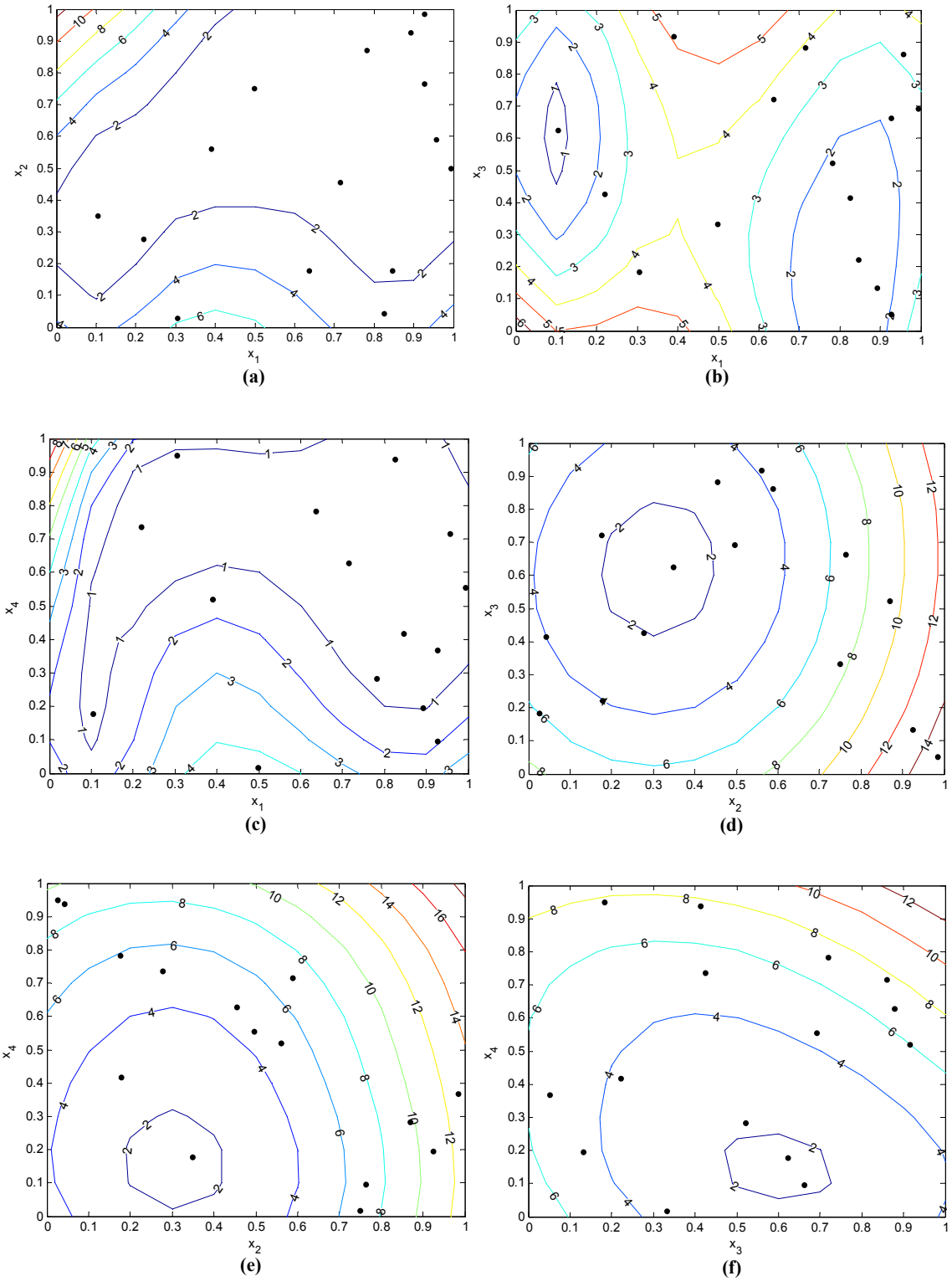


Fig. 8 MSE plots for pair-wise design variables after the 2nd stage

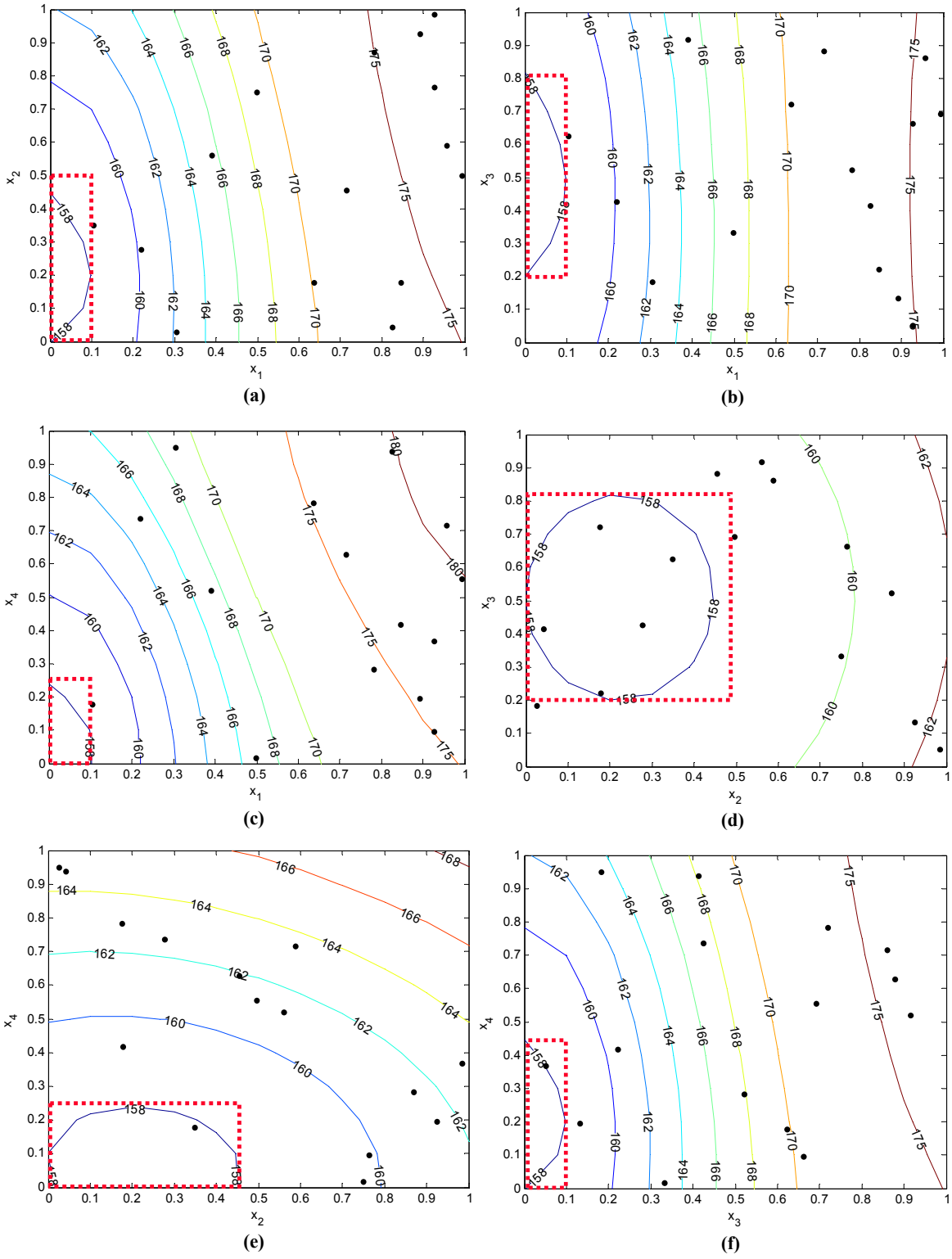


Fig. 9 Pair-wise interaction plots from the 2nd stage model

With the 6 additional design points, a new Kriging model was developed using the total of 15 points. The updated model parameters are $\hat{\sigma}^2 = 27.2614$, $\hat{\boldsymbol{\theta}} = (4.1054, 0.5849, 0.3153, 0.8876)$, and $\hat{\beta} = 171.82$. The prediction of the minimum from the second stage model is $\hat{y}$= 157.5744, corresponding to the same design parameters as in Table 2 and the true response value of y=154.44. Obviously, the minimum point has not changed but we have successfully reduced the prediction uncertainty to be much less than 5% in majority of the regions, as evidenced by the MSE plots in Fig. 8. With this small level of uncertainty, the minimal value region indicated by the prediction model can be deemed trustworthy.

3.3.2. Refine the location of the minimum point -- the 3rd stage

The 2nd stage Kriging model predicts that the minimum response value is about y=154.44, which sets a target for our zoom-in effort. Also noticing that among the first 15 design points the minimum y is 158.62, we decide to set the sub-region boundary to be the contour line of 158. We select the sub-region by using interaction contour plots between a pair of variables, as shown in Fig. 9. In those graphic plots, we are slightly conservative in drawing the region so that the 158 line is included.

As done earlier we get the zoom-in region from Fig. 9 by taking the union of sub-regions from each of the contour plots. The following region is obtained as the region to be zoomed in.

$$0 \le x_1 \le 0.1; \quad 0 \le x_2 \le 0.5; \quad 0 \le x_3 \le 0.8; \quad 0 \le x_4 \le 0.5 \qquad (13)$$

We again choose a 6-point LHS design over this region (the design matrix is also in the Appendix) and concatenate the data points to our previous design to get a modified 21-point design. A new 3rd stage Kriging model was fitted using these 21 points and the updated model parameters are $\hat{\sigma}^2 = 56.89$, $\hat{\boldsymbol{\theta}} = (2.2029, 0.1449, 0.8688, 4.7316)$, and $\hat{\beta} = 167.83$. This Kriging model finds a smaller value of the response function and the corresponding design variables are different. The results from the 3rd stage model are given in Table 3.

Table 3 Results after the 3rd step

x_1	x_2	x_3	x_4	y	$\hat{y}$
20.0	12.0	2.2	16.5	153.47	154.31

3.4 Final Step and Stopping Rule

The "zooming-in" procedure as described above will usually be continued repeatedly until the model so obtained is sufficiently accurate and the area of interest (e.g., a minimum region) is located. We will use the MSE plot to characterize the model accuracy and use the change in the minimum response values in two consecutive steps to benchmark if an area of interest is located. If the increment/decrement in the response values is not significant and the MSE plot shows no large uncertainty value, then we can stop the sequential strategy.

For the problem in hand, the decrease in the minimum response values found by the 3rd stage model is only 1% from the 2nd stage model. The largest MSE value is only about 2% (the MSE plot for x_1, x_2 is shown in Fig. 10 (b)). Both criteria are satisfied and we can conclude that an adequate Kriging model is established to replace the FEA model in the numerical integration in equation (2). Therefore, we choose to stop at this stage.

The contour plots of response function and the MSE value for the final Kriging model are shown in Fig. 10 (only the plots for x_1, x_2 are shown). The Kriging contour provides a global view of the response surface over the design space; the whole response surface has a decreasing trend along its reverse diagonal direction and the minimum region can be easily located to be the left-bottom corner. Please recall that the Kriging model we used in this example only has a constant polynomial term β_1. However, it appears that we do not have to include a first order term because the resulting Kriging model (mainly through its correlation term Z) is able to capture the surface change over the design space adequately.

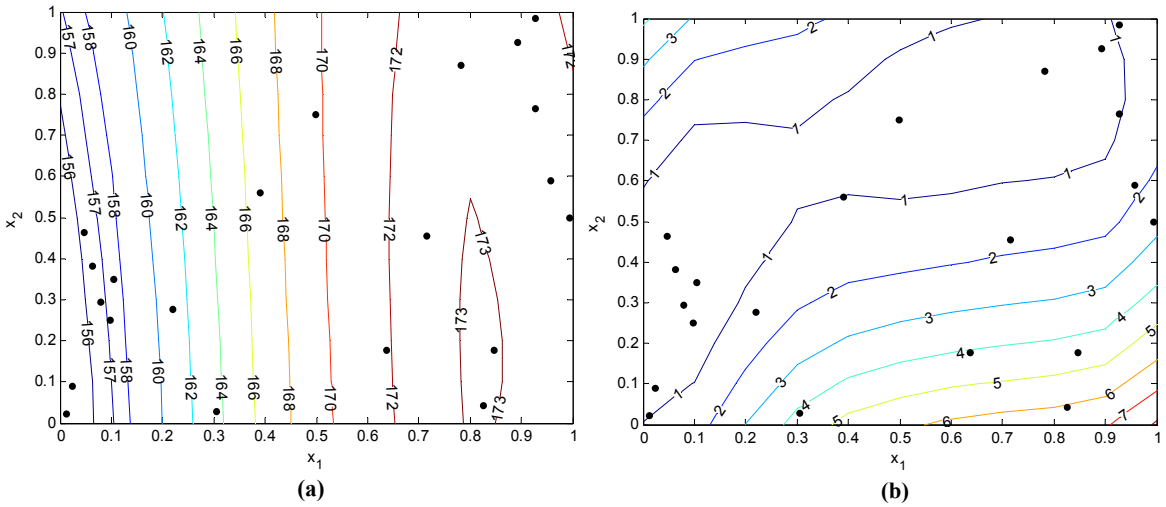


Fig. 10 Contour plots: (a) response surface and (b) MSE value for the 21-point model after 3rd stage

In order to carry out the numerical integration in equation (2), we also need to determine the distribution of noise factors. They will usually be assumed normal or uniform based on engineering understanding of the design variables, or empirically determined using historical data, or a combination of both. We here choose to omit the numerical integration part because it will follow a straightforward procedure after we find a computationally simple Kriging model to surrogate the FEA model (for a numerical integration procedure, please refer to [16] or use *MATLAB* quadrature functions).

CHAPTER IV

COMPARISON WITH CLASSICAL DESIGNS

Traditional response surface methodology also follows a sequential strategy. It usually starts with a first order fractional factorial design, then finds the probing direction based on the steepest decent method, and finally switches to a second order second (such as a Central Composite Design) to locate the optimal point when it approaches an optimum (likely a local one) [17,18].

We employ the traditional RSM to the electronic packaging problem. Because of the lack of prior knowledge of where to start our search, we simply choose the starting point as the current operating conditions, i.e., the center point of our design space. A linear model was fitted around this starting point. Because the assumption of linearity will hold over a small region, we will limit the first step design to a smaller hypercube whose edge size is $1/4^{th}$ the size of the design range. This choice of $1/4^{th}$ is empirical since there is usually no established rule in literature. Given four design parameters, we use a 2^{4-1} fractional factorial design, which will guarantee that the main effects can be reasonably estimated since the aliased three-factor interactions are usually not significant. From this experiment, a first order equation is obtained, which is given below.

$$\hat{y} = 118.3 + 1.4788x_1 + 0.122085x_2 - 0.0955x_3 + 0.442x_4 \qquad (14)$$

The coefficients of the first order model indicate the steepest direction. Following the direction, we will change the settings of design variables by taking a few sequential steps. Since the coefficient of x_1 is the greatest, we can see that the greatest decrease will occur if we change the values with respect to x_1. If we take the step sizes relative to x_1 (when coded a unit step corresponds to 1.875 of x_1) then, the steepest descent vector is (-1.875, -0.3963, + 0.0086, - 0.5604), which will be added to the current design setting sequentially. For each step, we get the corresponding simulation value, as shown in Fig. 11 (a), which gives us an idea of the validity. We go along this direction unless we

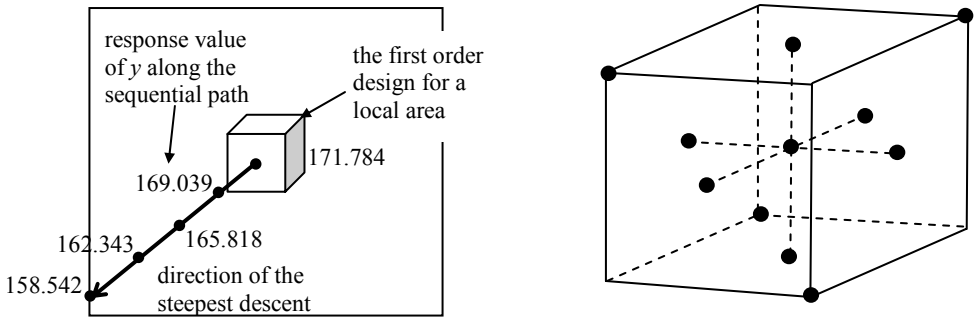
either hit the boundary of the design space or the simulation values start increasing instead of decreasing (i.e. we encounter a turning point). In our case after five steps, we reach the design space boundary.

Then, we fit a second order model around this point using a Small Composite Design [18], one of the second order design methods similar to Face Centered Design and its design settings for three variables are illustrated in Fig. 11 (b). This Small Composite Design is based on a 2^{4-1} fractional factorial but includes eight more design points at the centers of each hyper-surface, and one design point at the center of the hypercube (Fig. 11 (b) is an illustration for 3 design variables). The edge size of the hypercube for the Small Composite Design is also chosen as the $1/4^{th}$ the size of the design range.

The fitted second order model is expressed as

$$\hat{y} = 66.85 + 4.83x_1 + 1.0x_2 + 1.267x_3 + 0.0466x_4 - 0.0067x_1x_2 - 0.0012x_1x_3 + 0.0242x_1x_4 - 0.03091x_2x_3 - 0.017x_2x_4 - 0.0236x_3x_4 - 0.0738x_1^2 - 0.004x_2^2 + 0.0059x_3^2 + 0.0054x_4^2 \quad (15)$$

Based on the above model, we can find the minimum value of the response function and the corresponding design parameters; they are shown in Table 4.



(a) Steepest Decent

(b) Small Composite Design

Fig. 11 Traditional response surface methodology for the electronic packaging problem

Table 4 Results from the steepest descent method

x_1	x_2	x_3	x_4	y	$\hat{y}$
20.0	12.0	1.0344	17.2416	153.40	151.50

Compared with the minimum response value found by the Kriging model, the difference is negligible, meaning both methods can successfully find the optimal region. However, the sequential Kriging method, as presented in this thesis, takes 22 runs (1st stage: 9; 2nd stage: 6; 3rd stage: 6; and the final solution: 1). The traditional RSM, on the other hand, requires 31 runs (8 runs from a 2^{4-1} design, 5 runs of steepest descent, 17 runs from the Face Centered Design and 1 run at the final solution). The number of computer experiments required by the sequential Kriging model is 41% less than the traditional RSM, which is a remarkable reduction.

In this electronic packaging example, the response surface of the Von Mises stress is actually not very complicated -- it has a global descending trend toward its left-bottom corner and it has only one minimum point. That is the reason that the traditional RSM gives us the same optimum values as the Kriging model does. Should we have a complicated surface as in Fig. 4, the chance that RSM locates the global optimum successfully is in fact not high. However, the sequential strategy presented in this thesis will not be bounded by the complexity of the surface due to the space filling design it used at each stage.

The advantages of the proposed sequential strategy also manifest in other aspects. One is that the final Kriging model will provide a global view of the response surface over the design space. By contrast, the traditional RSM will fall short of doing so since it fits polynomial models to a few small areas. It will be difficult to piece together the global response surface.

Another advantage is the uncertainty evaluation. The Kriging model provides an uncertainty evaluation of response prediction over the entire design space. For the traditional RSM, uncertainty can be evaluated based on linear regression theory.

However, with the absence of random errors, the uncertainty evaluation from linear regression is not guaranteed to be the interpolation uncertainty in a computer experiment. Once again, those uncertainty evaluations are only available for a few small areas in the design space. A high uncertainty could exist elsewhere without being detected simply because the steepest descent method does not lead the experimenter there.

CHAPTER V

CONCLUDING REMARKS

A systematic approach for design optimization using the Kriging model was employed in this thesis. A sequential strategy, which adds extra design points in the region where more information would be needed, is developed to update the design of computer simulations and revise the subsequent Kriging models for prediction. The sequential strategy demonstrates a superior performance by finding the minimum Von Mises stress in an electronic packaging process while using 40% less computer runs compared to a classical RSM design approach. It also provides the global view of the response surface and the uncertainty level over the entire region, of which the traditional RSM falls short.

Despite the current advancements in design and analysis of computer experiments, we would also like to point out the challenges ahead of us. One is the way we are deciding the zoom-in sub-regions. Currently we are dependent on a visual technique that assesses the interaction contour plots to gain information. This visual technique is intuitive and easy to use for practitioners. However, it can become cumbersome when the design space has a high dimension. It would be desirable to develop a searching procedure to locate the area of interest in a high dimension design space.

That naturally leads us to the issues of screening out the important factors. In this study, our collaborating industrial partner has already identified the four design parameters for investigation. In other circumstances, however, screening itself could be an issue, especially considering that a computer experiment often has many more factors and could possibly take multiple levels. The screening task could be more challenging than that in physical experiments.

The final note is on the space-filling design, specifically the LHS design. The LHS design does not possess the nice properties such as orthogonality of factorial designs.

Therefore, it is possible that the LHS design results in an ill-conditioned dataset, where $\mathbf{R}_D$ has some near-zero eigenvalues. The orthogonal array-based LHS [19] could help overcome this problem. However, it also causes a significant increase in experiment runs.

REFERENCES

[1] Darveaux, R., 2000, "Effect of Simulation Methodology on Solder Joint Crack Growth Correlation," *Proceedings of the 50th IEEE Electronic Components and Technology Conference* (ECTC'00), Las Vegas, NV, May 21 - 24.

[2] Shetty, S., and Reinikainen, T., 2001, "Three- and Four-Point Bend Testing for Electronic Packages," *Proceedings of the 2001 EuroSimE,* Paris, France, April 9 - 11 .

[3] Rodgers, B., Punch, J., Jarvis, J., Myllykoski, P., and Reinikainen, T., 2002, "Finite Element Modeling of a BGA Package Subjected to Thermal and Power Cycling," *Inter Society Conference on Thermal Phenomena in Electronic Systems,* San Diego, CA, May 30 – June 1.

[4] ANSYS, Inc., 2002, *ANSYS 6.1 Documentation*, ANSYS, Inc., Canonsburg, PA.

[5] Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments (With Discussion)," Statistical Science, **4**, pp. 409-435.

[6] Sacks, J., Schiller, S. B., and Welch, W. J., 1989, "Design for Computer Experiments," Technometrics, **31**, pp. 41-47.

[7] Welch, W.J., Buck, R.J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M.D., 1992, "Screening, Predicting and Computer Experiments," Technometrics, **34**, pp. 15-25.

[8] Ripley, B. D., 1981, *Spatial Statistics,* John Wiley & Sons, New York

[9] Bendat, J.S., and Piersol, A.G., 1986, *Random Data: Analysis and Measurement Procedures,* 2nd ed. (revised and expanded), John Wiley & Sons, New York, NY.

[10] Santner, T. J., Williams, B.J., Notz, W.I., 2003, *Design and Analysis of Computer Experiments,* Springer Series in Statistics, Springer Verlag, New York, NY.

[11] Lindstrom, M.J. and Bates, D.M. 1988, "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," Journal of the American Statistical Association, **83**, pp. 1014 -1022.

[12] Nelder, J.A. and Mead, R., 1965. "A Simplex Method for Function Minimization," The Computer Journal, **7**, pp. 308-313.

[13] Bernardo, M. C., Buck, R., Liu, L., Nazaret, W. A., Sacks, J., and Welch, W. J., 1992, "Integrated Circuit Design Optimization Using a Sequential Strategy", IEEE Transactions in Computer-Aided Design, **11**, pp. 361-372.

[14] Chang, P. B., Williams, B. J., Bhalla, K. S. B., Belknap, T. W., Santner, T. J., Notz, W.I., Bartel, D. L., 2001, "Design and Analysis of Robust Total Joint Replacements: Finite Element Model Experiments with Environmental Variables", Journal of Biomechanical Engineering, **123**, pp. 239-246.

[15] McKay, M. D., Beckman, R. J., and Conover, W. J., 1979, "A Comparison of Three Methods for Selecting Values of Input Variables In the Analysis of Output from a Computer Code", Technometrics, **21**, pp. 239--245.

[16] Davis, P. J. and Rabinowitz, P., 1984, *Methods of Numerical Integration*, Second Edition, Academic Press, New York, NY.

[17] Montgomery, D. C., 1997, *Design and Analysis of Experiments*, Fourth Edition, John Wiley & Sons, New York, NY.

[18] Myers, R. H., and Montgomery, D. C., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley and Sons, New York, NY.

[19] Owen, A.B., 1992, "Orthogonal Arrays for Computer Experiments, Integration and Visualization", Statistica Sinica, **2**, pp. 439-452.

APPENDIX

Table 5 Design table for the sequential experiments in Chapter III

	Runs	x_1	x_2	x_3	x_4	y
1st stage	1	21.58	20.37	3.50	17.64	158.62
	2	25.85	25.44	4.66	22.79	169.37
	3	30.74	22.91	4.52	24.39	176.75
	4	23.31	18.65	2.70	26.03	165.25
	5	33.90	35.65	1.21	20.52	180.24
	6	31.75	32.90	3.08	19.22	176.60
	7	29.56	16.25	3.88	26.74	175.45
	8	24.56	12.63	1.73	29.21	167.88
	9	27.48	29.99	2.33	15.24	168.63
2nd stage	10	34.93	23.94	3.77	23.29	181.38
	11	32.39	13.013	2.65	29.05	179.08
	12	33.40	34.22	1.53	17.92	178.09
	13	32.69	16.27	1.89	21.24	176.49
	14	34.36	26.14	4.44	25.73	182.02
	15	33.90	30.33	3.65	16.43	177.14
3rd stage	16	21.47	17.97	3.22	19.15	158.78
	17	20.20	12.51	2.32	22.03	156.82
	18	20.70	23.11	1.43	16.46	156.61
	19	20.35	14.14	1.71	18.12	155.59
	20	21.18	19.07	3.75	20.36	158.89
	21	20.95	21.16	3.12	15.71	156.35

VITA

Abhishek Gupta was born on September 30, 1980 in Bangalore, India. In 2002, he received his Bachelor of Technology from the Indian Institute of Technology, Delhi. He joined the Industrial Engineering Department at Texas A&M University in Fall 2002. He was awarded the 2004 Mary G. Natrella Scholarship from the Quality and Productivity Research section of the American Statistical Association. He received his Master of Science degree in Industrial Engineering in May 2004.

Permanent address is:
Flat 303, Kunal-3
Plots 250-251-252
Sector 21, Nerul
Navi Mumbai - 400706
INDIA
Phone: (91)-22-27719369
E-mail: abhishek.gupta@tamu.edu
abhigp@yahoo.com